Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
ELSEVIER
Journal of Multivariate Analysis 92 (2005) 386–404
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

Jackknifing in partially linear regression models with serially correlated errors

Jinhong You,[a,*] Xian Zhou,[a] and Gemai Chen[b]

[a] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China
[b] Department of Mathematics & Statistics, University of Calgary, Calgary, Alberta, T2N 1N4 Canada

Received 5 December 2002

Abstract

In this paper jackknifing technique is examined for functions of the parametric component in a partially linear regression model with serially correlated errors. By deleting partial residuals a jackknife-type estimator is proposed. It is shown that the jackknife-type estimator and the usual semiparametric least-squares estimator (SLSE) are asymptotically equivalent. However, simulation shows that the former has smaller biases than the latter when the sample size is small or moderate. Moreover, since the errors are correlated, both the Tukey type and the delta type jackknife asymptotic variance estimators are not consistent. By introducing cross-product terms, a consistent estimator of the jackknife asymptotic variance is constructed and shown to be robust against heterogeneity of the error variances. In addition, simulation results show that confidence interval estimation based on the proposed jackknife estimator has better coverage probability than that based on the SLSE, even though the latter uses the information of the error structure, while the former does not.

AMS 2000 subject classifications: primary 62G05; secondary 62G20, 62M10

Keywords: Partially linear regression model; Serially correlated errors; Semiparametric least-squares estimator; Jackknife-type estimation; Asymptotic normality; Consistency; Robustness

*Corresponding author. Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA.
E-mail address: jinhongyou@hotmail.com (J. You).

doi:10.1016/j.jmva.2003.11.004

1. Introduction

Regression analysis is one of the most mature and widely applied branches of statistics. For a long time, its main theory is on either parametric or nonparametric regressions. Recently, however, semiparametric regressions have attracted increasing attention from statisticians and practitioners using statistics. One main reason, we think, is that semiparametric regression can reduce the high risk of model misspecification relative to a fully parametric model on the one hand, and avoid some serious drawbacks of pure nonparametric methods on the other hand. Of importance in the class of semiparametric regression models is the *partially linear regression* model proposed by Engle et al. [4] in a study of the effects of weather on electricity demand. A partially linear regression model has the form

$$y_i = x_i'\beta + g(t_i) + \varepsilon_i, i = 1, \dots, n, \tag{1.1}$$

where y_i's are responses, $x_i = (x_{i1}, \dots, x_{ip})'$ and $t_i \in [0, 1]$ are design points, $\beta = (\beta_1, \dots, \beta_p)'$ is an unknown parameter vector, $g(\cdot)$ is an unknown bounded real-valued function defined on $[0, 1]$, ε_i's are unobservable random errors, and the prime $'$ denotes the transpose of a vector or matrix.

Model (1.1) has been widely studied in the literature; see, for example, the work of Heckman [10], Rice [16], Chen [1], Speckman [23], Robinson [17], Chen and Shiau [2], Donald and Newey [3], Eubank and Speckman [5], Gao [7], Hamilton and Truong [8], Shi and Li [22], and Liang et al. [13], among others. More references and techniques can be found in the recent monograph by Härdle et al. [9].

Fueled by modern computing power, jackknife, bootstrap and other resampling methods are extensively used in many statistical applications due to their strong capability of approximating unknown probability distributions and then its characteristics like moments or confidence regions for unknown parameters. There has, however, been little work on applying resampling methods to semiparametric regressions in the literature, apart from Liang et al. [14] and You and Chen [26]. Liang et al. [14] considered bootstrap estimation of β and the error variance in model (1.1) with i.i.d. errors. You and Chen [26] proposed a jackknife-type estimator for a function of β in model (1.1) with i.i.d. errors. Different from the usual jackknife estimation by deleting original data points, You and Chen's procedure relies on deleting partial residuals. There are two motivations for applying jackknife method. One is that jackknife estimation can reduce estimation bias and here almost all estimators are biased in semiparametric regressions. The other is that it can provide consistent estimators of the asymptotic covariance matrices, and according to Shi and Lau [21], this is not an easy job in semiparametric regressions.

In some applications, the independence assumption on the errors is not appropriate. For example, in the study of the effect of weather on electricity demand, Engle et al. [4] found that the data were autocorrelated at order one. Similar to the traditional jackknife method, the jackknife-type estimation proposed in You and Chen [26] did not take into account the fact that the data may be serially correlated. This results in inconsistency of the jackknife variance estimator. In order to accommodate the serial correlation in the data, we introduce cross-product terms in the process of constructing a jackknife asymptotic variance estimator. We show that the resulted estimator is consistent and robust against heterogeneity of the error variances. Simulation results show that the jackknife-type estimators for β or functions of β have smaller biases than the usual semiparametric least-squares estimators (SLSE) when the sample size is small or moderate. Moreover, confidence interval estimation based on the jackknife-type estimator has better coverage probability than that based on the SLSE, even though the latter uses the information of the error structure, while the former does not.

The rest of this paper is organized as follows. The usual SLSE is presented in Section 2 together with our assumptions. Our jackknife-type estimation procedure is discussed in Section 3. A consistent jackknife-type estimator of the asymptotic variance is given in Section 4. Some simulation results are reported in Section 5, followed by concluding remarks in Section 6. Several technical lemmas are relegated to the appendix.

2. Preliminary

Throughout this paper we assume that the errors $\{\varepsilon_i\}$ in model (1.1) is an MA(∞) process, namely,

$$\varepsilon_i = \sum_{j=0}^{\infty} \phi_j e_{i-j}, \quad \text{with} \quad \sum_{j=0}^{\infty} |\phi_j| < \infty, \tag{2.1}$$

where $e_j, j = 0, \pm 1, \ldots$ are i.i.d. random variables with $Ee_j = 0$ and Var $(e_j) = \sigma_e^2 < \infty$. We also assume that $1_n = (1, \ldots, 1)'$ is not in the space spanned by the column vectors of $X = (x_1, \ldots, x_n)'$, which ensures the identifiability of the model in (1.1) according to Chen [1]. Moreover, suppose, as is common in the setting of partially linear regression model, that $\{x_i\}$ and $\{t_i\}$ are related via $x_{is} = h_s(t_i) + u_{is}$, $i = 1, \ldots, n,\ s = 1, \ldots, p$. The reasonableness of this relation can be found in Speckman [23].

In order to construct our jackknife-type estimator, we need an estimator for β first. For convenience we here adopt the partial kernel smoothing method proposed by Speckman [23] to construct an SLSE of β, although the jackknife methodology proposed here would be equally applicable to other estimations methods.

Suppose that $\{x_i', t_i, y_i; i = 1, \ldots, n\}$ satisfy model (1.1). The SLSE constructed using the partial kernel smoothing method of Speckman [23] is of the form

$$\widehat{\beta}_n = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'\widehat{y}, \tag{2.2}$$

where $\widehat{y} = (\widehat{y}_1, \ldots, \widehat{y}_n)'$, $\widehat{X} = (\widehat{x}_1, \ldots, \widehat{x}_n)'$, $\widehat{y}_i = y_i - \sum_{j=1}^n W_{nj}(t_i)y_j$, $\widehat{x}_i = x_i - \sum_{j=1}^n W_{nj}(t_i)x_j$ for $i = 1, \ldots, n$, and $W_{nj}(\cdot)$ are weight functions defined below. In some applications the parameter of interest is $\theta = m(\beta)$, where $m(\cdot)$ is a function from $\mathscr{R}^p$ to $\mathscr{R}$. Such parameter functions include, for example, the roots and turning points of the mean polynomials in polynomial regression models. A natural estimator of $m(\beta)$ is $\widehat{\theta}_n = m(\widehat{\beta}_n)$.

In order to study the asymptotic properties of $\widehat{\beta}_n$ and $\widehat{\theta}_n$, we make the following assumptions. These assumptions, while look a bit lengthy, are actually quite mild and can be easily satisfied (see Remarks 2.1–2.4 following the assumptions).

Assumption 2.1. The probability weight functions $W_{ni}(\cdot)$ satisfy

(i) $\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n} W_{ni}(t_j) = O(1)$,
(ii) $\max_{1\leqslant i,j\leqslant n} W_{ni}(t_j) = O(n^{-2/3})$,
(iii) $\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n} W_{ni}(t_j)I(|t_i - t_j| > c_n) = O(d_n)$, where $I(A)$ is the indicator function of a set A, c_n satisfies $\limsup_{n\to\infty} nc_n^3 < \infty$ and d_n satisfies $\limsup_{n\to\infty} nd_n^3 < \infty$.

Assumption 2.2. $\operatorname{mineig}(n^{-1}\sum_{i=1}^{n} u_i u_i')$ is bounded away from 0, $||u_i|| \leqslant c$, $i = 1, \dots, n$, and $\max_{1\leqslant i\leqslant n} ||\sum_{j=1}^{n} W_{nj}(t_i)u_j|| = o[n^{-\frac{1}{6}}(\log n)^{-1}]$, where $||\cdot||$ denotes the Euclidean norm, c is a positive constant and $\operatorname{mineig}(\cdot)$ represents the minimum eigenvalue of a symmetric matrix.

Assumption 2.3. The functions $g(\cdot)$ and $h_s(\cdot)$ satisfy the Lipschitz condition of order 1 on $[0, 1]$ for $s = 1, \dots, p$.

Assumption 2.4. The spectral density function $\psi(\cdot)$ of $\{\varepsilon_i\}$ is bounded away from 0 and ∞. In addition, $\sup_n n\sum_{j=n}^{\infty} |\phi_j| < \infty$ and $Ee_0^4 < \infty$.

Remark 2.1. For $0 = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n \leqslant t_{n+1} = 1$ such that $\max_{1\leqslant i\leqslant n+1} |t_i - t_{i-1}| = O(n^{-1})$, take $W_{nj}(\cdot)$ as the k_n nearest neighbor type weight functions, namely

$$W_{nj}(t) = \begin{cases} k_n^{-1} & \text{if } t_j \text{ belongs to the } k_n \text{ nearest neighbor of } t, \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1, \dots, n$ and $k_n = n^{2/3}$. Then $W_{nj}(\cdot)$ satisfy Assumption 2.1.

Remark 2.2. Note that $\sum_{j=1}^{n} W_{nj}(t_i)u_j$ is a weighted average of the locally centered quantities $\{u_j\}_{j=1}^{n}$. Therefore, $\max_{1\leqslant i\leqslant n} ||\sum_{j=1}^{n} W_{nj}(t_i)u_j|| = o[n^{-\frac{1}{6}}(\log n)^{-1}]$ is a mild condition. The condition of $\operatorname{mineig}(n^{-1}\sum_{i=1}^{n} u_i u_i')$ being bounded away from 0 is necessary when we derive asymptotical distributions for various estimators.

Remark 2.3. Assumption 2.3 is mild and holds for most commonly used functions, such as the polynomial and trigonometric functions.

Remark 2.4. Obviously, the usual finite parameter AR, MA or ARMA processes satisfy Assumption 2.4.

Theorem 2.1. *Suppose that Assumptions* 2.1–2.4 *hold. Let* $\widehat{\beta}_n$ *be defined in* (2.2). *Then*

$$\Sigma^{-\frac{1}{2}}(\widehat{\beta}_n - \beta) \rightarrow_{\mathrm{D}} N(0, I_{\mathrm{p}}) \quad \text{as } n \rightarrow \infty,$$

where "$\rightarrow_{\mathrm{D}}$" *denotes convergence in distribution,* $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$, $\Sigma^{1/2}$ *is a square root of* $\Sigma = (U'U)^{-1}U'\Omega U(U'U)^{-1}$, $U = (u_1, \ldots, u_n)'$, *and* $\Omega = (\gamma_\varepsilon(i-j))_{i,j=1}^n$ *with* $\gamma_\varepsilon(h)$ *denoting the autocovariance function of* $\{\varepsilon_i\}$ *at* lag h.

A proof of Theorem 2.1 can be found in You [25]. Applying the delta method, $\widehat{\theta}_n$ is also asymptotically normally distributed. Its asymptotic variance is

$$V_\delta = (\nabla m(\beta))'(U'U)^{-1}U'\Omega U(U'U)^{-1}\nabla m(\beta),$$

where $\nabla m(\beta)$ is the gradient of $m(\cdot)$ at β. Based on $\widehat{\beta}_n$ we obtain estimated residuals $\widehat{\varepsilon}_i = \widehat{y}_i - \widehat{x}_i'\widehat{\beta}_n$, $i = 1, \ldots, n$. Then a conventional estimator of V_δ is

$$\widehat{V}_\delta = (\nabla m(\widehat{\beta}_n))'(\widehat{X}'\widehat{X})^{-1}\widehat{X}'\widehat{\Omega}\widehat{X}(\widehat{X}'\widehat{X})^{-1}\nabla m(\widehat{\beta}_n),$$

where $\widehat{\Omega}$ is the $n \times n$ matrix with $\widehat{\varepsilon}_i\widehat{\varepsilon}_j$ as its (i,j) element. This estimator, however, is not consistent for V_δ at order n^{-1} [11], i.e., $n(\widehat{V}_\delta - V_\delta)$ does not converge to zero in probability as $n \rightarrow \infty$. Therefore, $\widehat{V}_\delta$ cannot be used to make asymptotically valid statistical inference. The simulation results in Section 5 confirm this point.

3. Jackknife-type estimator for the parametric component function

In this and the following sections we construct jackknife-type estimators for $\theta = m(\beta)$ and its asymptotic variance. Due to the existence of the nonparametric component $g(\cdot)$, it is difficult to use the traditional jackknife method directly. So instead of deleting the original data points, we delete partial residuals $(\widehat{x}_i, \widehat{y}_i)$ to construct jackknife-type estimators. Once the partial residuals $(\widehat{x}_i, \widehat{y}_i)$ are computed, the computation of the jackknife-type estimators for the partially linear regression model (1.1) follows the same procedure as that of the linear regression model; see [19].

Let $\widehat{\beta}_{n,-i}$ be an estimator similar to $\widehat{\beta}_n$ except that $\widehat{X}$ and $\widehat{y}$ are replaced by $\widehat{X}_{-i}$ and $\widehat{y}_{-i}$ where $\widehat{X}_{-i}$ is obtained by deleting the ith row $\widehat{x}_i$ from the matrix $\widehat{X}$ and $\widehat{y}_{-i}$ is obtained by deleting the ith element $\widehat{y}_i$ from the vector $\widehat{y}$. Then $\widehat{\beta}_{n,-i}$ has the following form:

$$\widehat{\beta}_{n,-i} = (\widehat{X}_{-i}'\widehat{X}_{-i})^{-1}\widehat{X}_{-i}'\widehat{y}_{-i}.$$

The ith pseudovalue is defined to be

$$\widehat{p}_i = m(\widehat{\beta}_n) - (n-1)[m(\widehat{\beta}_{n,-i}) - m(\widehat{\beta}_n)] \quad \text{for } i = 1, \ldots, n, \tag{3.1}$$

and a jackknife-type estimator for θ is the mean of these pseudovalues, namely,

$$\widehat{J}_n = \frac{1}{n}\sum_{i=1}^{n} \widehat{p}_i = m(\widehat{\beta}_n) - \frac{n-1}{n}\sum_{i=1}^{n} [m(\widehat{\beta}_{n,-i}) - m(\widehat{\beta}_n)]. \tag{3.2}$$

The following theorem shows that $\widehat{J}_n$ and $\widehat{\theta}_n$ are asymptotically equivalent.

Theorem 3.1. *Suppose that $\theta = m(\beta)$ where $m(\cdot)$ has bounded second partial derivatives in some neighborhood of β. If Assumptions* 2.1–2.4 *hold then we have*

$$V_\delta^{-\frac{1}{2}}(\widehat{\theta}_n - \theta) - V_\delta^{-\frac{1}{2}}(\widehat{J}_n - \theta) \rightarrow_{\mathrm{p}} 0 \quad \text{as } n \rightarrow \infty,$$

where "$\rightarrow_{\mathrm{p}}$" denotes convergence in probability and V_δ is defined in the last section.

Remark 3.1. Applying Theorems 2.1 and 3.1 and the delta method, it follows that $V_\delta^{-1/2}(\widehat{J}_n - \theta)$ converges in distribution to $N(0,1)$ as $n \rightarrow \infty$.

Proof of Theorem 3.1. Let $S(\beta, s)$ be the sphere with radius s and center β. From Theorem 2.1, we have $P\{\widehat{\beta}_n \in S(\beta, s/2)\} \rightarrow 1$ as $n \rightarrow \infty$, and it follows from Lemma A.5 in the appendix that $P\{\widehat{\beta}_{n,-i} \in S(\beta, s/2), i = 1, \ldots, n\} \rightarrow 1$ as $n \rightarrow \infty$. Therefore, imposing or removing the condition that $C_n = \{\widehat{\beta}_n, \widehat{\beta}_{n,-1}, \ldots, \widehat{\beta}_{n,-n} \in S(\beta, s)\}$ holds has no effect on any limiting probabilities. For $\widehat{\beta}_n, \widehat{\beta}_{n,-1}, \ldots, \widehat{\beta}_{n,-n} \in S(\beta, s)$, by Taylor's expansion we have

$$\begin{aligned}\widehat{J}_n &= m(\widehat{\beta}_n) - \frac{n-1}{n}\left[\nabla m(\widehat{\beta}_n)\right]' \sum_{i=1}^{n} (\widehat{\beta}_{n,-i} - \widehat{\beta}_n) \\ &\quad - \frac{n-1}{2n}\sum_{i=1}^{n} (\widehat{\beta}_{n,-i} - \widehat{\beta}_n)' \nabla^2 m(\xi_i)(\widehat{\beta}_{n,-i} - \widehat{\beta}_n),\end{aligned}$$

where $\nabla^2 m(\cdot)$ is the second partial derivatives of $m(\cdot)$, ξ_i is a point on the line segment between $\widehat{\beta}_n$ and $\widehat{\beta}_{n,-i}$. As a result, in order to prove Theorem 3.1, it suffices to show that the second and third terms on the right-hand side of the above equation are $o_{\mathrm{p}}(n^{1/2})$. To prove that this is true for the second term, note that similar to Lemma 3.2 of Miller [15] we have

$$\widehat{\beta}_n - \widehat{\beta}_{n,-i} = \frac{(\widehat{X}'\widehat{X})^{-1}\widehat{x}_i(\widehat{y}_i - \widehat{x}_i'\widehat{\beta}_n)}{1 - \widehat{x}_i'(\widehat{X}'\widehat{X})^{-1}\widehat{x}_i} = (1 - w_i)^{-1}(\widehat{X}'\widehat{X})^{-1}\widehat{x}_i(\widehat{y}_i - \widehat{x}_i'\widehat{\beta}_n), \tag{3.3}$$

where $w_i = \widehat{x}_i'(\widehat{X}'\widehat{X})^{-1}\widehat{x}_i$. Therefore, combining $\sum_{i=1}^{n} \widehat{x}_i\widehat{\varepsilon}_i = 0$, it holds that

$$n^{\frac{1}{2}}\sum_{i=1}^{n} (\widehat{\beta}_{n,-i} - \widehat{\beta}_n) = -n^{\frac{1}{2}}(\widehat{X}'\widehat{X})^{-1}\sum_{i=1}^{n} \frac{\widehat{x}_i\widehat{\varepsilon}_i}{1 - w_i} = -n^{\frac{1}{2}}(\widehat{X}'\widehat{X})^{-1}\sum_{i=1}^{n} \frac{w_i\widehat{x}_i\widehat{\varepsilon}_i}{1 - w_i}.$$

By Assumptions 2.1–2.3, Lemma A.1 and the fact

$$(A+aB)^{-1} = A^{-1} - aA^{-1}BA^{-1} + O(a^2),$$

we have $(\widehat{X}'\widehat{X})^{-1} = O(n^{-1})$. Therefore, it is sufficient to show that

$$\sum_{i=1}^{n} n^{-\frac{1}{2}}(1-w_i)^{-1} w_i \widehat{x}_i \widehat{\varepsilon}_i \to_{\mathrm{p}} 0 \quad \text{as } n \to \infty. \tag{3.4}$$

Note that by Assumptions 2.1–2.3, $\max_{1 \leqslant i \leqslant n} w_i = O(n^{-1})$ and

$$E(\widehat{x}_i \widehat{\varepsilon}_i) = \sum_{i=1}^{n} \widehat{x}_{is} \left[\tilde{g}(t_i) - \widehat{x}_i'(\widehat{X}'\widehat{X})^{-1} \sum_{j=1}^{n} \widehat{x}_j \tilde{g}(t_j) \right] = o(n^{\frac{1}{2}}),$$

where $\tilde{g}(t_i) = g(t_i) - \sum_{j=1}^{n} W_{nj}(t_i) g(t_j)$. Therefore, the expectation of the summation in (3.4) tends to zero. On the other hand,

$$\begin{aligned}\max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} |\gamma_\varepsilon(i-j)| &\leqslant 2 \sum_{h=0}^{\infty} |\gamma_\varepsilon(h)| \leqslant 2\sigma_e^2 \sum_{k=0}^{\infty} |\phi_k| \sum_{h=0}^{\infty} |\phi_{k+h}| \leqslant 2\sigma_e^2 \left(\sum_{k=0}^{\infty} |\phi_k| \right)^2 \\ &= O(1).\end{aligned}$$

This implies $\operatorname{maxeig}(\Omega) = O(1)$, where $\Omega = \operatorname{Cov}(\varepsilon)$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$, and $\operatorname{maxeig}(\cdot)$ denotes the maximum eigenvalue of a symmetric matrix. So the variance of the sth coordinate of $\sum_{i=1}^{n} (w_i \widehat{x}_i \widehat{\varepsilon}_i)/(n^{1/2}(1-w_i))$ is

$$\begin{aligned}
&\operatorname{Var}\left[\frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^{n} \frac{w_i \widehat{x}_{is} \widehat{\varepsilon}_i}{(1-w_i)} \right] \\
&\quad = \frac{1}{n} \left(\frac{w_1 \widehat{x}_{1s}}{1-w_1}, \ldots, \frac{w_n \widehat{x}_{ns}}{1-w_n} \right) \left[I - \widehat{X}(\widehat{X}'\widehat{X})^{-1}\widehat{X}' \right] \\
&\qquad \times (I-W)\Omega(I-W)'[I - \widehat{X}(\widehat{X}'\widehat{X})^{-1}\widehat{X}'] \left(\frac{w_1 \widehat{x}_{1s}}{1-w_1}, \ldots, \frac{w_n \widehat{x}_{ns}}{1-w_n} \right)' \\
&\quad \leqslant \operatorname{maxeig}(\Omega) \cdot \frac{2}{n} \left[1 + \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ni}^2(t_j) \right] \sum_{i=1}^{n} \left(\frac{w_i}{1-w_i} \right)^2 \widehat{x}_{is}^2 \\
&\quad = O(n^{-1}) \cdot \left(1 + n \max_{1 \leqslant i,j \leqslant n} W_{ni}(t_j) \right) \cdot \max_{1 \leqslant i \leqslant n} \frac{w_i^2}{(1-w_i)^2} \cdot \sum_{i=1}^{n} \widehat{x}_{is}^2.
\end{aligned}$$

Since $\max_{1 \leqslant i \leqslant n} w_i^2/(1-w_i)^2 \to 0$ and $\sum_{i=1}^{n} \widehat{x}_{is}^2/n$ is bounded, the above variance converges to zero. Hence (3.4) holds.

By the assumption that the second partial derivatives of $m(\cdot)$ are bounded and Lemma A.4, it easy to see that the third term in the decomposition of $\widehat{J}_n$ is $o_{\mathrm{p}}(n^{-\frac{1}{2}})$. The proof is complete. □

Remark 3.2. Shao [18] showed that $(n-1)n^{-1}\sum_{i=1}^{n}[m(\widehat{\beta}_{n,-i})-m(\widehat{\beta}_n)]$ can be used to approximate the bias of $m(\widehat{\beta}_n)$ in the case of ordinary linear regression model because $\widehat{\beta}_n$ is unbiased. This approach, however, does not seem to be applicable in our case since $\widehat{\beta}_n$ defined in (2.2) is biased.

4. A consistent estimator for the asymptotic variance

When the errors are independent, Tukey [24] proposed to use the sample variance of the pseudovalues

$$\frac{1}{n-1}\sum_{i=1}^{n}(\widehat{p}_i-\widehat{J}_n)^2$$

divided by n to estimate the asymptotic variance of $\widehat{J}_n$, where $\widehat{p}_i$ are defined in (3.1). Miller [15] and You and Chen [26] showed that this estimator is consistent in the context of linear regression models and partially linear regression models, respectively. However, when the errors are dependent, this estimator is not consistent because it does not take into account the correlated error structure.

Motivated by Lele [12], we propose an estimator of the asymptotic variance of $\widehat{J}_n$ of the form

$$\widehat{V}_J=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}(\widehat{p}_i-\widehat{J}_n)(\widehat{p}_j-\widehat{J}_n),$$

where m_n is a positive integer dependent on n. It should be noted that this type of estimator was also discussed in Shao and Tu [19, p. 392].

The following theorem shows that $\widehat{V}_J$ is a consistent estimator of the asymptotic variance of $\widehat{J}_n$.

Theorem 4.1. *Let $\theta=m(\beta)$, where $m(\cdot)$ has continuous second partial derivatives in some neighborhood of β. If Assumptions* 2.1–2.4 *hold, $m_n\to\infty$ and $m_n^2/n\to 0$* as $n\to\infty$, *then we have*

$$n\widehat{V}_J-nV_\delta\to_p 0\quad\text{as } n\to\infty.$$

Proof. The proof is similar to that of Theorem 2 of Miller [15]. For $\widehat{\beta}_n,\widehat{\beta}_{n-1},\dots,\widehat{\beta}_{n-n}\in S(\beta,s)$, $m(\widehat{\beta}_{n,-i})=m(\widehat{\beta}_n)+(\widehat{\beta}_{n,-i}-\widehat{\beta}_n)'\nabla m(\xi_i)$, where ξ_i is a point on the line segment between $\widehat{\beta}_n$ and $\widehat{\beta}_{n,-i}$. For $i=1,\dots,n$, let

$$U_i=\widehat{\varepsilon}_i\widehat{x}_i'(\widehat{X}'\widehat{X})^{-1}[\nabla m(\xi_i)-\nabla m(\widehat{\beta}_n)]\quad\text{and}\quad V_i=\frac{w_i}{1-w_i}\widehat{\varepsilon}_i\widehat{x}_i'(\widehat{X}'\widehat{X})^{-1}\nabla m(\xi_i).$$

Then, as $\sum_{i=1}^{n} \widehat{x}_i' \widehat{\varepsilon}_i = 0$, it holds that

$$\sum_{i=1}^{n} \sum_{|j-i| \leqslant m_n} (\widehat{p}_i - \widehat{J}_n)(\widehat{p}_j - \widehat{J}_n)$$
$$= (n-1)^2 \sum_{i=1}^{n} \sum_{|j-i| \leqslant m_n} \{\widehat{\varepsilon}_i \widehat{x}_i' (\widehat{X}'\widehat{X})^{-1} [\nabla m(\widehat{\beta}_n)]' + U_i - \overline{U} + V_i - \overline{V}\}$$
$$\times \{\widehat{\varepsilon}_j \widehat{x}_j' (\widehat{X}'\widehat{X})^{-1} [\nabla m(\widehat{\beta}_n)]' + U_j - \overline{U} + V_j - \overline{V}\},$$

where $\overline{U} = \sum_{i=1}^{n} U_i/n$ and $\overline{V} = \sum_{i=1}^{n} V_i/n$. From Lemmas A.4 and A.5 and the root-n consistency of $\widehat{\beta}_n$, it follows that

$$(\nabla m(\widehat{\beta}_n))'(n-1)(\widehat{X}'\widehat{X})^{-1} \left[\sum_{i=1}^{n} \sum_{|j-i| \leqslant m_n} \widehat{\varepsilon}_i \widehat{\varepsilon}_j \widehat{x}_i \widehat{x}_j' \right] (\widehat{X}'\widehat{X})^{-1} \nabla m(\widehat{\beta}_n) - n V_\delta = o_{\mathrm{p}}(1).$$

Since

$$\max_{1 \leqslant i \leqslant n} [\nabla m(\xi_i) - \nabla m(\widehat{\beta}_n)] = o_{\mathrm{p}}(1) \quad \text{and} \quad \max_{1 \leqslant i \leqslant n} w_i = o(1),$$

we have

$$(n-1)^2 \sum_{i=1}^{n} \sum_{|j-i| \leqslant m_n} (U_i - \overline{U})(U_j - \overline{U}) = o_{\mathrm{p}}(n)$$

and

$$(n-1)^2 \sum_{i=1}^{n} \sum_{|j-i| \leqslant m_n} (V_i - \overline{V})(V_j - \overline{V}) = o_{\mathrm{p}}(n).$$

Moreover, by Cauchy–Schwarz inequality, the three cross-product sums (divided by $n-1$) will also converge to zero, and the proof is complete. □

Shao and Wu [20] proved that the jackknife variance estimator was robust against heteroscedasticity in conventional linear regression models. We show that this robustness property remains true when the errors are serially correlated and the regression model is semiparametric. For the partially linear regression model (1.1), if the errors are $\sigma_i \varepsilon_i$, where σ_i are positive constants and ε_i are defined in (2.2) with variance 1, then the following Theorem 4.2 states that under certain regularity conditions, the estimator $\widehat{V}_J$ is still a consistent estimator of the true variance of $\widehat{\theta}_n$.

Theorem 4.2. *Suppose that the errors in model* (1.1) *are $\sigma_i\varepsilon_i$, where σ_i^2 are bounded away from* 0 *and* ∞, *and Assumptions* 2.1 *to* 2.4 *hold. If $m(\cdot)$ has continuous second partial derivatives in some neighborhood of β, $m_n \to \infty$ and $m_n^2/n \to 0$ as $n \to \infty$, then we have*

$$\begin{aligned} &n\widehat{V}_J - n(\nabla m(\beta))'(U'U)^{-1}U' \operatorname{diag}(\sigma_1, \dots, \sigma_n)\Omega \operatorname{diag}(\sigma_1, \dots, \sigma_n)U(U'U)^{-1}\nabla m(\beta) \\ &\quad = o_{\mathrm{p}}(1). \end{aligned}$$

Proof. A simple modification of Lemma A.6 enables us to show that

$$\sum_{i=1}^{n} \sum_{|j-i| \leqslant m_n} \widehat{\varepsilon}_i\widehat{\varepsilon}_j\widehat{x}_i\widehat{x}_j' - U' \operatorname{diag}(\sigma_1, \dots, \sigma_n)\Omega \operatorname{diag}(\sigma_1, \dots, \sigma_n)U \to_{\mathrm{p}} 0 \quad \text{as } n \to \infty.$$

Using this result the theorem follows from similar arguments to those used in the proof of Theorem 4.1. □

We can apply the asymptotic results of Theorems 3.1, 4.1 and 4.2 to construct tests and confidence intervals for the parametric component β.

Corollary 4.1. *If the conditions of Theorem* 4.1 *are satisfied, then under the null hypothesis H_0: $m(\beta) = 0$,*

$$n\widehat{J}_n^2 \widehat{V}_J^{-1} \to_{\mathrm{D}} \chi_1^2 \quad \text{as } n \to \infty,$$

where χ_1^2 is the chi-square distribution with one degree of freedom.

5. A simulation study

From Theorem 3.1 we know that the jackknife-type estimator $\widehat{J}_n$ and the SLSE $\widehat{\theta}_n$ are asymptotically equivalent. Here, we conduct a simulation study to compare their performances in terms of bias, mean square error and coverage probability for finite samples.

Example 5.1. The observations are generated from

$$y_i = 3.5x_i + \sin(4\pi t_i) + 2.5\varepsilon_i, \quad i = 1, \dots, n,$$

where $t_i = (i - 0.5)/n$ and x_i, ε_i are generated as follows:

- $x_i = 5t_i^2 + \eta_i$, η_i are i.i.d. $U(-0.5, 0.5)$,
- $\{\varepsilon_i\}$ is an AR(1) process: $\varepsilon_i = 0.5\varepsilon_{i-1} + e_i$, $i = 1, \dots, n$, where e_i are i.i.d. $U(-0.5, 0.5)$.

For a given sample size n, we generate 10,000 samples from the above model (the x_i values are generated once for a given n) and estimate β and $\sqrt{\beta}$ for each sample by SLSE and jackknife estimations. We perform the smoothing with different nearest-neighbor parameter k_n using a grid search. It turns out that the results for the parametric component (or its function) and asymptotic variance are insensitive to the choice of the nearest-neighbor parameter k_n. Biases and variances of the above estimators are listed in Table 1.

We also simulate interval estimations. Altogether five methods labeled I, II, III, IV and V are compared. Methods II and III utilize the information of the error structure, and method III serves as a benchmark. A full description of the five methods is given below.

Method	Approximate $100(1-\alpha)\%$ confidence intervals
I	$\widehat{\beta}_n \pm z_{\alpha/2}\sqrt{\widehat{V}_\delta}$, where $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the $N(0,1)$.
II	$\widehat{\beta}_n \pm z_{\alpha/2}\sqrt{\widehat{V}_1}$ with $\widehat{V}_1 = \sigma_\varepsilon^2(\widehat{X}'\widehat{X})^{-1}\widehat{X}'(\widehat{\rho}^{\lvert i-j\rvert})\widehat{X}(\widehat{X}'\widehat{X})^{-1}$, where $\sigma_\varepsilon^2 = n^{-1}\sum_{i=1}^n \widehat{\varepsilon}_i^2$ and $\widehat{\rho} = (\sum_{i=2}\widehat{\varepsilon}_i\widehat{\varepsilon}_{i-1})/\sum_{i=1}^n \widehat{\varepsilon}_i^2$.
III	$\widehat{\beta}_n \pm z_{\alpha/2}\sqrt{\widehat{V}_1}$ with $\widehat{V}_1 = \sigma_\varepsilon^2(\widehat{X}'\widehat{X})^{-1}\widehat{X}'(\rho^{\lvert i-j\rvert})\widehat{X}(\widehat{X}'\widehat{X})^{-1}$, where σ_ε^2 and ρ are known.
IV	$\widehat{J}_n \pm z_{\alpha/2}\sqrt{\widehat{V}_J}$ with $m_n = [n^{0.2}]$, where $[n^{0.2}]$ denotes the integer part of $n^{0.2}$.
V	$\widehat{J}_n \pm z_{\alpha/2}\sqrt{\widehat{V}_J}$ with $m_n = [n^{0.4}]$.

The coverage probabilities and the median lengths of the confidence intervals are presented in Table 2.

Example 5.1 (Continued). We also consider a moving average process for the errors. Assume that $\{\varepsilon_i\}$ is an MA(1) process specified by

$$\varepsilon_i = 0.5e_i + e_{i-1}, \quad i = 1, \ldots, n.$$

Simulation results parallel to Tables 1 and 2 are listed in Tables 3 and 4 below.

From Tables 1–4 we make the following observations:

1. From Tables 1 and 3 we can see that the jackknife-type estimators of the parametric component or its function perform better than the SLSEs in terms of biase and variance, especially when the sample size n is small ($n = 20, 30, 40$). As n increases, the difference diminishes.
2. Tables 1 and 3 also show that both the jackknife-type estimators and the SLSEs underestimate the parametric component or its function. A more detailed study to find out the reason seems worthwhile.

Table 1
Biases and variances of the SLSE and the jackknife-type estimator with $\beta = 3.5$ and AR(1) errors

	$\theta = \beta$				$\theta = \sqrt{\beta}$			
n	Bias $(\widehat{\theta}_n)$	Var $(\widehat{\theta}_n)$	Bias $(\widehat{J}_n)$	Var $(\widehat{J}_n)$	Bias $(\widehat{\theta}_n)$	Var $(\widehat{\theta}_n)$	Bias $(\widehat{J}_n)$	Var $(\widehat{J}_n)$
20	−0.5377	0.2892	−0.3838	0.1473	−0.1626	0.0264	−0.0835	0.0069
30	−0.4691	0.2201	−0.2895	0.0838	−0.1204	0.0145	−0.0634	0.0040
40	−0.3942	0.1554	−0.2392	0.0572	−0.1189	0.0142	−0.0464	0.0022
50	−0.3654	0.1336	−0.1737	0.0302	−0.1003	0.0101	−0.0621	0.0039
75	−0.3820	0.1459	−0.2632	0.0695	−0.0899	0.0081	−0.0634	0.0040
100	−0.3521	0.1243	−0.2869	0.0823	−0.0886	0.0078	−0.0722	0.0052
200	−0.2821	0.0796	−0.2401	0.0576	−0.0749	0.0056	−0.0654	0.0043

Table 2
Coverage probabilities and median lengths (in parenthese) of the five confidence intervals with nominal confidence level 0.95, $\beta = 3.5$ and AR(1) errors

n	I	II	III	IV	V
20	0.02% (0.0001)	69.5% (0.6719)	98.3% (1.4639)	79.8% (1.0495)	80.2% (1.0968)
30	0.01% (0.0001)	75.2% (0.5180)	99.8% (1.3279)	85.1% (0.8022)	82.4% (0.7264)
40	0.03% (0.0002)	76.2% (0.5584)	99.8% (1.5043)	88.5% (0.9489)	87.2% (0.9400)
50	0.07% (0.0002)	69.6% (0.4578)	99.7% (1.2513)	90.5% (0.8822)	87.8% (0.8434)
75	0.05% (0.0002)	64.7% (0.3445)	99.2% (0.9494)	89.5% (0.7003)	86.8% (0.6728)
100	0.07% (0.0001)	59.2% (0.2556)	97.6% (0.7035)	91.8% (0.6027)	88.9% (0.5778)
200	0.04% (0.0002)	55.6% (0.2154)	96.4% (0.5981)	93.5% (0.5605)	91.6% (0.5445)

Table 3
Biases and variances of the SLSE and the jackknife-type estimator with $\beta = 3.5$ and MA(1) errors

	$\theta = \beta$				$\theta = \sqrt{\beta}$			
n	Bias $(\widehat{\theta}_n)$	Var $(\widehat{\theta}_n)$	Bias $(\widehat{J}_n)$	Var $(\widehat{J}_n)$	Bias $(\widehat{\theta}_n)$	Var $(\widehat{\theta}_n)$	Bias $(\widehat{J}_n)$	Var $(\widehat{J}_n)$
20	−0.3519	0.1239	−0.1303	0.0169	−0.1217	0.0148	−0.0821	0.0067
30	−0.4001	0.1607	−0.2055	0.0422	−0.0972	0.0094	−0.0662	0.0044
40	−0.3110	0.0967	−0.1615	0.0261	−0.0885	0.0078	−0.0150	0.0023
50	−0.3415	0.1166	−0.1850	0.0342	−0.0960	0.0092	−0.0809	0.0065
75	−0.3113	0.0969	−0.2302	0.0588	−0.0836	0.0070	−0.0653	0.0043
100	−0.3397	0.1154	−0.3073	0.0945	−0.0739	0.0055	−0.059	0.0035
200	−0.3016	0.0909	−0.2840	0.0806	−0.0683	0.0046	−0.0602	0.0036

3. In Tables 2 and 4, the converge probability of confidence interval I is very low, almost equal to zero. We think the reason is the inconsistency of the asymptotic variance estimator $\widehat{V}_\delta$.

Table 4
Coverage probabilities and median lengths (in parenthese) of the five confidence intervals with nominal confidence level 0.95, $\beta = 3.5$ and MA(1) errors

n	I	II	III	IV	V
20	0.00% (0.0001)	41.4% (0.5852)	79.2% (1.2787)	89.5% (1.9603)	89.1% (2.0687)
30	0.04% (0.0001)	67.7% (0.4634)	99.4% (1.1898)	86.3% (0.8836)	85.6% (0.8711)
40	0.03% (0.0001)	57.8% (0.4712)	97.1% (1.2651)	85.2% (1.0789)	84.6% (1.0953)
50	0.03% (0.0001)	79.6% (0.4743)	98.9% (0.9152)	84.7% (0.6368)	81.8% (0.5860)
75	0.07% (0.0002)	61.6% (0.3394)	98.6% (0.9406)	90.5% (0.7576)	87.6% (0.7313)
100	0.00% (0.0001)	58.4% (0.2579)	97.7% (0.7179)	91.7% (0.6214)	88.6% (0.5893)
200	0.05% (0.0002)	65.2% (0.2081)	95.7% (0.5853)	93.0% (0.5707)	90.8% (0.5501)

4. Confidence intervals IV and V do not use the information of the error structure, but they still perform better than interval II which uses the information of the error structure. We think the reason is that the performance of the estimator $\widehat{\rho}$ is not good when the sample size is small or moderate.
5. Jackknife intervals IV and V have very similar performances for the two choices of m_n.
6. When sample size n increases, the performance of interval IV and interval V become closer to that of the benchmark.

Remark 5.1. From Tables 1 and 3 we can see that even for the case of $m(x) = x$ (linear estimator), the jackknife estimator still out-performs the original SLSE. This is not the case for ordinary linear regression models. We think the reason is that $\widehat{\beta}_n$ defined in (2.2) is biased because of the non-parametric component $g(\cdot)$, while $\widehat{\beta}_n$ in ordinary linear regression models is unbiased. To some extent, this implies that the advantage of the jackknife method is even greater in semiparametric regression models than in ordinary linear regression models.

6. Concluding remarks

In this paper we have studied the estimation of partially linear regression models with serially correlated errors. We proposed jackknife-type estimators for the parametric component or its functions, and showed that they are asymptotically equivalent to the SLSE and are asymptotically normally distributed. We also constructed a jackknife-type consistent estimator of the asymptotic variance. Simulation results show that our jackknife-type estimators perform better than the SLSEs when sample size is small or moderate.

In the process of constructing the jackknife asymptotic variance estimator we need a truncation number m_n. Is there an optimal choice of m_n under $m_n \to \infty$ and

$m_n^2/n \to 0$ as $n \to \infty$? Will the choice of m_n depend upon specific smoothing method, error structure or covariates? These are interesting problems that deserve further research efforts.

Appendix A

In this appendix we present several lemmas which are used to prove the main results in Sections 3 and 4. Lemma A.1 below is trivial to prove.

Lemma A.1. *Suppose that Assumptions* 2.1 *and* 2.3 *hold. Then as* $n \to \infty$,

$$\max_{0 \leqslant s \leqslant p} \max_{1 \leqslant i \leqslant n} \left| G_s(t_i) - \sum_{j=1}^{n} W_{nj}(t_i) G_s(t_j) \right| = O(c_n) + O(d_n),$$

where $G_0(\cdot) = g(\cdot)$ *and* $G_s(\cdot) = h_s(\cdot),\ s = 1, \ldots, p$.

The next lemma is a generalization of Lemma A.3 in Härdle et al. [9] under i.i.d. assumption to the case of a linear process.

Lemma A.2. *Suppose that* $\{\varepsilon_i\}$ *is a linear process defined by* (2.1). *If Assumptions* 2.2 *and* 2.4 *hold, then as* $n \to \infty$,

$$\max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^{n} W_{nj}(t_i) \varepsilon_j \right| = O(n^{-\frac{1}{3}} \log n) \quad a.s.$$

Proof. Following a procedure widely used in time series to truncate the MA(∞) error process $\{\varepsilon_i\}$ into two parts, and similar to the proof of Lemma A.3 in Härdle et al. [19], we can prove this lemma. The details can be found in You [25].

Lemma A.3. *Suppose that Assumptions* 2.1–2.4 *hold. Let* $\widehat{g}_n(\cdot) = \sum_{j=1}^{n} W_{nj}(\cdot)(y_j - x_j'\widehat{\beta}_n)$. *Then we have*

$$\max_{1 \leqslant i \leqslant n} |\widehat{g}_n(t_i) - g(t_i)| = O_{\mathrm{p}}(n^{-\frac{1}{3}} \log n).$$

Proof. By the root-n consistency of $\widehat{\beta}_n$ and Lemmas A.1 and A.2, it is easy to complete the proof. $\square$

Lemma A.4. *Suppose that Assumptions* 2.1–2.4 *hold. Then we have*

$$\sum_{i=1}^{n} \widehat{\varepsilon}_i^2 \widehat{x}_i \widehat{x}_i' = O_{\mathrm{p}}(n), \quad \text{where} \quad \widehat{\varepsilon}_i = \widehat{y}_i - \widehat{x}_i' \widehat{\beta}_n \text{ for } i = 1, \ldots, n.$$

Proof. From the definition of $\widehat{\varepsilon}_i$, it holds that $\widehat{\varepsilon}_i = x_i'(\beta - \widehat{\beta}_n) + (g(t_i) - \widehat{g}_n(t_i)) + \varepsilon_i$. Therefore,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \widehat{\varepsilon}_i^2 \widehat{x}_i \widehat{x}_i' =& \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \widehat{x}_i \widehat{x}_i' + \frac{2}{n}\sum_{i=1}^n \varepsilon_i \Big[x_i'(\beta - \widehat{\beta}_n) + (g(t_i) - \widehat{g}_n(t_i))\Big]\widehat{x}_i \widehat{x}_i' \\ &+ \frac{1}{n}\sum_{i=1}^n \Big[x_i'(\beta - \widehat{\beta}_n) + (g(t_i) - \widehat{g}_n(t_i))\Big]^2 \widehat{x}_i \widehat{x}_i' \\ =& I_1 + I_2 + I_3, \quad \text{say.}\end{aligned}$$

By Cauchy–Schwarz inequality, it suffices to show that, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \widehat{x}_i \widehat{x}_i' = O_{\mathrm{p}}(1) \tag{A.1}$$

and

$$\frac{1}{n}\sum_{i=1}^n [x_i'(\beta - \widehat{\beta}_n) + (g(t_i) - \widehat{g}_n(t_i))]^2 \widehat{x}_i \widehat{x}_i' = O_{\mathrm{p}}(1). \tag{A.2}$$

It is easy to see that

$$\left|\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \widehat{x}_{is_1}\widehat{x}_{is_2} - \frac{\sigma_\varepsilon^2}{n}\sum_{i=1}^n \widehat{x}_{is_1}\widehat{x}_{is_2}\right| \leqslant \left[\frac{1}{n}\sum_{i=1}^n (\varepsilon_i^2 - \sigma_\varepsilon^2)\widehat{x}_{is_1}^2\right]^{1/2}\left[\frac{1}{n}\sum_{i=1}^n (\varepsilon_i^2 - \sigma_\varepsilon^2)\widehat{x}_{is_2}^2\right]^{\frac{1}{2}},$$

where $\sigma_\varepsilon^2 = \mathrm{Var}\ (\varepsilon_i)$. Moreover, since $\widehat{x}_{is}$ is uniformly bounded for all $i = 1, \ldots, n$ and $s = 1, \ldots, p$, we have

$$\begin{aligned}& E\left[\frac{1}{n}\sum_{i=1}^n (\varepsilon_i^2 - \sigma_\varepsilon^2) x_{is_1}^2\right]^2 \\ &\quad = \frac{Ee_0^4 - 3\sigma_e^4}{n^2}\sum_{i_1=1}^n \sum_{i_2=1}^n \widehat{x}_{i_1 s_1}^2 \widehat{x}_{i_2 s_1}^2 \left[\sum_{j=0}^{\infty} \phi_j^2 \phi_{j+i_2-i_1}^2 + 2\gamma_\varepsilon(i_2 - i_1)^2\right] = o(1).\end{aligned}$$

Therefore, (A.1) holds. Furthermore, by the root-n consistency of $\widehat{\beta}_n$ and Lemma A.3 we have

$$\begin{aligned}& \frac{1}{n}\sum_{i=1}^n [x_i'(\beta - \widehat{\beta}_n) + (g(t_i) - \widehat{g}_n(t_i)]^2 \widehat{x}_{is}^2 \\ &\quad \leqslant \left\{\max_{1\leqslant i\leqslant n}[x_i'(\beta - \widehat{\beta}_n) + (g(t_i) - \widehat{g}_n(t_i))]^2\right\}\frac{1}{n}\sum_{i=1}^n \widehat{x}_{is}^2 \\ &\quad \leqslant 2\left[\max_{1\leqslant i\leqslant n} x_i'(\beta - \widehat{\beta}_n)(\beta - \widehat{\beta}_n)' x_i + \max_{1\leqslant i\leqslant n} (g(t_i) - \widehat{g}_n(t_i))^2\right]\frac{1}{n}\sum_{i=1}^n \widehat{x}_{is}^2 \\ &\quad = o_{\mathrm{p}}(1).\end{aligned}$$

This proves (A.2) and completes the proof. $\square$

Define $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)'$, $\tilde{\varepsilon}_i = \sum_{j=1}^n W_{nj}(t_i)\varepsilon_j$, $\tilde{G} = (\tilde{g}(t_1), \ldots, \tilde{g}(t_n))'$ and

$$\tilde{g}(t_i) = g(t_i) - \sum_{j=1}^n W_{nj}(t_i)g(t_j).$$

Lemma A.5. *Suppose that Assumptions* 2.1–2.4 *hold. Then it holds that*

$$P\{(\widehat{\beta}_n - \widehat{\beta}_{n,-i})'(\widehat{\beta}_n - \widehat{\beta}_{n,-i}) \leqslant \delta, i = 1, \ldots, n\} \to 1 \quad \textit{as } n \to \infty \textit{ for any } \delta > 0.$$

Proof. By (3.3) and Chebyshev inequality it can be seen that

$$\begin{aligned} &P\left\{\max_{1 \leqslant i \leqslant n} (\widehat{\beta}_n - \widehat{\beta}_{n,-i})'(\widehat{\beta}_n - \widehat{\beta}_{n,-i}) > \delta\right\} \\ &\quad \leqslant \sum_{i=1}^n P\left\{\widehat{\varepsilon}_i^2 > \frac{\delta(1-w_i)^2}{\widehat{x}_i'(\widehat{X}'\widehat{X})^{-2}\widehat{x}_i}\right\} \leqslant \sum_{i=1}^n \frac{\widehat{x}_i'(\widehat{X}'\widehat{X})^{-2}\widehat{x}_i}{\delta(1-w_i)^2} E\widehat{\varepsilon}_i^2. \end{aligned}$$

Now we prove $E\widehat{\varepsilon}_i^2 = O(1)$. By the definition of $\widehat{\varepsilon}_i$,

$$\begin{aligned} E\widehat{\varepsilon}_i^2 &= E\left[\varepsilon_i - \sum_{j=1}^n W_{nj}(t_i)\varepsilon_j + \tilde{g}(t_i) - \widehat{x}_i'(\beta - \widehat{\beta}_n)\right]^2 \\ &\leqslant \sigma_\varepsilon^2 + O(n^{-\frac{2}{3}}) + E\left[\sum_{j=1}^n W_{nj}(t_i)\varepsilon_j\right]^2 + E\{\widehat{x}_i'(\widehat{X}'\widehat{X})^{-1}\widehat{X}'(\varepsilon - \tilde{\varepsilon})(\varepsilon - \tilde{\varepsilon})'\widehat{X}(\widehat{X}'\widehat{X})^{-1}\widehat{x}_i\} \\ &\quad + \widehat{x}_i'(\widehat{X}'\widehat{X})^{-1}\widehat{X}'\tilde{G}\tilde{G}'\widehat{X}(\widehat{X}'\widehat{X})^{-1}\widehat{x}_i. \end{aligned}$$

Obviously, $E\varepsilon_{j_1}\varepsilon_{j_2} = \sum_{k=0}^\infty \phi_k\phi_{k+|j_2-j_1|}\sigma_e^2$. Therefore,

$$\begin{aligned} E\left[\sum_{j=1}^n W_{nj}(t_i)\varepsilon_j\right]^2 &\leqslant \left[n \max_{1 \leqslant i,j \leqslant n} W_{nj}^2(t_i)\right] \sum_{h=-(n-1)}^{n-1} (n-|h|)\sigma_e^2 \sum_{k=0}^\infty |\phi_k\phi_{k+|h|}| \\ &\leqslant O(n^{-\frac{1}{3}})\left[\sum_{k=0}^\infty \phi_k^2\sigma_e^2 + 2\sigma_e^2\sum_{h=1}^{n-1}\sum_{k=0}^\infty |\phi_k\phi_{k+|h|}|\right] \\ &= O(n^{-\frac{1}{3}}) = o(1). \end{aligned}$$

Moreover, the (s_1, s_2) element of $\widehat{X}'\tilde{G}\tilde{G}'\widehat{X}$ satisfies $(\sum_{j=1}^n \tilde{g}(t_i)\widehat{x}_{is_1})(\sum_{j=1}^n \tilde{g}(t_i)\widehat{x}_{is_2}) = o(n^2)$ by Assumptions 2.1 and 2.2 and Lemma A.1. Combining

$$\max_{1 \leqslant i \leqslant n} \widehat{x}_i'\widehat{x}_i = O(1) \quad \text{and } (\widehat{X}'\widehat{X})^{-1} = O(n^{-1}), \tag{A.3}$$

it follows that the last term in the decomposition of $E\widehat{\varepsilon}_i^2$ is $o(1)$ for $i = 1, \ldots, n$. Similarly, we can show that the third term in the decomposition of $E\widehat{\varepsilon}_i^2$ is $O(1)$. So $E\widehat{\varepsilon}_i^2 = O(1)$ for $i = 1, \ldots, n$. Then by (A.3) we get

$$P\left\{\max_{1\leqslant i\leqslant n}(\widehat{\beta}_n - \widehat{\beta}_{n,-i})'(\widehat{\beta}_n - \widehat{\beta}_{n,-i}) > \delta\right\} \to 0 \text{ as } n \to \infty \text{ for any } \delta,$$

and the proof is complete. □

Lemma A.6. *Suppose that Assumptions* 2.1–2.4 *hold. If* $m_n \to \infty$ *and* $m_n^2/n \to 0$ *as* $n \to \infty$, *then*

$$\sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}\widehat{\varepsilon}_i\widehat{\varepsilon}_j\widehat{x}_i\widehat{x}_j' - U'\Omega U = o_{\mathrm{p}}(n),$$

where U *and* Ω *are defined in Theorem* 2.1.

Proof. By the definition of $\widehat{\varepsilon}_i$, we have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}\widehat{\varepsilon}_i\widehat{\varepsilon}_j\widehat{x}_i\widehat{x}_j' &= \sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}\varepsilon_i\varepsilon_j\widehat{x}_i\widehat{x}_j' + \sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}\tilde{\varepsilon}_i\tilde{\varepsilon}_j\widehat{x}_i\widehat{x}_j' \\ &+ \sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}\tilde{g}(t_i)\tilde{g}(t_j)\widehat{x}_i\widehat{x}_j' + \sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}\widehat{x}_i'(\beta-\widehat{\beta}_n)\widehat{x}_j'(\beta-\widehat{\beta}_n)\widehat{x}_i\widehat{x}_j' \\ &- 2\sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}[\varepsilon_i\tilde{\varepsilon}_j - \varepsilon_i\tilde{g}(t_j) + \tilde{\varepsilon}_i\tilde{g}(t_j) - \varepsilon_i\widehat{x}_j'(\beta-\widehat{\beta}_n) + \tilde{\varepsilon}_i\widehat{x}_j'(\beta-\widehat{\beta}_n) \\ &- \tilde{g}(t_i)\widehat{x}_j'(\beta-\widehat{\beta}_n)] = I_1 + \cdots + I_5, \quad \text{say.}\end{aligned}$$

Applying Lemmas A.1 and A.2 and the boundedness of $\widehat{x}_{is}$, it is easy to show that $I_i = o_{\mathrm{p}}(n)$ for $i = 2, 3$. Similarly, by the root-n consistency of $\widehat{\beta}_n$, it holds that $I_4 = o_{\mathrm{p}}(n)$. By Cauchy–Schwarz inequality we just need to show $I_1 - U'\Omega U = o_{\mathrm{p}}(n)$. Let

$$\Delta_{s_1s_2} = \sum_{i=1}^{n}\sum_{|j-i|\leqslant m_n}\varepsilon_i\varepsilon_j\widehat{x}_{is_1}\widehat{x}_{js_2}' - \sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_\varepsilon(i-j)\widehat{x}_{is_1}\widehat{x}_{js_2}'.$$

We want to prove $\Delta_{s_1s_2} = o_{\mathrm{p}}(n)$. It is easy to see that

$$|E\Delta_{s_1s_2}| = \left|\sum_{i=1}^{n}\sum_{|j-i|>m_n}\widehat{x}_{is_1}\widehat{x}_{js_2}\gamma_\varepsilon(i-j)\right| = O(1)\sum_{i=1}^{n}\sum_{|j-i|>m_n}|\gamma_\varepsilon(i-j)| = o(n)$$

because $m_n \to \infty$ and $\sum_{h=0}^{\infty} |\gamma_\varepsilon(h)| < \infty$. Hence, to finish the proof, it is sufficient to show that Var $(\Delta_{s_1 s_2}) = o(n^2)$. By (6.2.5) in Fuller [6],

$$\begin{aligned}\text{Var } (\Delta_{s_1 s_2}) \leqslant O(1) \cdot \sum_{i_1=1}^{n} \sum_{|j_1 - i_1| \leqslant m_n} \sum_{i_2=1}^{n} \sum_{|j_2 - i_2| \leqslant m_n} \\ \times \Bigg[\Bigg| (Ee_0^4 - 3\sigma_e^4) \sum_{l=0}^{\infty} \phi_l \phi_{l+i_1-j_1} \phi_{l+i_2-j_1} \phi_{l+j_2-j_1} \Bigg| \\ + |\gamma_\varepsilon(i_2 - i_1)\gamma_\varepsilon(j_2 - j_1)| + |\gamma_\varepsilon(i_2 - j_1)\gamma_\varepsilon(j_2 - i_1)| \Bigg] . = I_1 + I_2 + I_3, \quad \text{say.}\end{aligned}$$

Since $\sum_{h=0}^{\infty} |\gamma_\varepsilon(h)| < \infty$, $\max_{0 \leqslant h < \infty} |\gamma_\varepsilon(h)|$ is bounded. This implies that

$$I_2 \leqslant O(m_n^2) \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} |\gamma_\varepsilon(i_2 - i_1)| = O(n m_n^2) = o(n^2).$$

Similarly, we can show that $I_3 = o(n^2)$. Moreover,

$$I_1 = \sum_{j_1=1}^{n} \sum_{l=0}^{\infty} |\phi_l| \sum_{i_1=1+l-j_1}^{n+l-j_1} |\phi_{i_1}| \sum_{i_2=1+l-j_1}^{n+l-j_1} |\phi_{i_2}| \sum_{j_2=1+l-j_1}^{n+l-j_1} |\phi_{j_2}| = O(n) = o(n^2),$$

and the proof follows. □

References

[1] H. Chen, Convergence rates for parametric components in a partially linear model, Ann. Statist. 16 (1988) 136–147.
[2] H. Chen, J. Shiau, Data-driven efficient estimation for a partially linear model, Ann. Statist. 22 (1994) 211–237.
[3] G. Donald, K. Newey, Series estimation of semilinear models, J. Multivariate Anal. 50 (1994) 30–40.
[4] R.F. Engle, W.J. Granger, J. Rice, A. Weiss, Semiparametric estimates of the relation between weather and electricity sales, J. Amer. Statist. Asso. 80 (1986) 310–319.
[5] R. Eubank, P. Speckman, Trigonometric series regression estimators with an application to partially linear models, J. Multivariate Anal. 32 (1990) 70–85.
[6] A. Fuller, Introduction to Statistical Time Series, Wiley, New York, 1976.
[7] J.T. Gao, Asymptotic theory for partially linear models, Comm. Statist. Theory Methods A24 (8) (1995) 1985–2009.
[8] A. Hamilton, K. Truong, Local linear estimation in partially linear models, J. Multivariate Anal. 60 (1997) 1–19.
[9] W. Härdle, H. Liang, J. Gao, Partially Linear Models, Physica-Verlag, Heidelberg, 2000.
[10] N. Heckman, Spline smoothing in a partially linear model, J. Roy. Statist. Soc. Ser. B 48 (1986) 244–248.
[11] R.W. Keener, J. Kmenta, C.N. Weber, Estimation of the covariance matrix of the least-squares regression coefficients when the disturbance covariance matrix is of unknown form, Econom. Theory 7 (1991) 22–45.
[12] S. Lele, Jackknifing linear estimating equations: asymptotic theory and applications in stochastic processes, J. Roy. Statist. Soc. Ser. B 53 (1991) 253–267.

[13] H. Liang, W. Härdle, R.J. Carroll, Estimation in a semiparametric partially linear errors-in-variables model, Ann. Statist. 27 (1999) 1519–1535.
[14] H. Liang, W. Härdle, V. Sommerfeld, Bootstrap approximation in a partially linear regression model, J. Statist. Plann. Infer. 91 (2000) 413–426.
[15] R.G. Miller JR., An unbalanced jackknife, Ann. Statist. 2 (1974) 880–891.
[16] J. Rice, Convergence rates for partially splined models, Statist. Probab. Lett. 4 (1986) 203–208.
[17] P. Robinson, Root-N-consistent semiparametric regression, Econometrica 56 (1988) 931–954.
[18] J. Shao, Bootstrap variance and bias estimation in linear models, Canad. J. Statist. 16 (1988) 371–381.
[19] J. Shao, D. Tu, The Jackknife and Bootstrap, Springer, New York, 1995.
[20] J. Shao, C.F.J. Wu, Heteroscedasticity-robustness of jackknife variance estimators in linear models, Ann. Statist. 15 (1987) 1563–1579.
[21] J. Shi, T.S. Lau, Empirical likelihood for partially linear models, J. Multivariate Anal. 72 (2000) 132–149.
[22] P. Shi, G. Li, A note of the convergence rates of M-estimates for partially linear model, Statistics 26 (1995) 27–47.
[23] P. Speckman, Kernel smoothing in partial linear models, J. Roy. Statist. Soc. Ser. B 50 (1988) 413–437.
[24] J. Tukey, Bias and confidence in not quite large samples, Ann. Math. Statist. 29 (1958) 614.
[25] J.H. You, Semiparametric regression models with serially correlated and/or heteroscedastic errors, Unpublished thesis, University of Regina, 2002.
[26] J. You, G. Chen, Jackknife type estimation for smooth functions of parametric component in partially linear models, Comm. Statist. Theory Methods, 2002, accepted for publication.